

Stochastic Models for Solar Power

Dimitra Politaki, Sara Alouf

▶ **To cite this version:**

Dimitra Politaki, Sara Alouf. Stochastic Models for Solar Power. EPEW 2017 - European Performance Engineering Workshop, Sep 2017, Berlin, Germany. pp.282–297, 10.1007/978-3-319-66583-2_18 . hal-01624419

HAL Id: hal-01624419
https://inria.hal.science/hal-01624419

Submitted on 26 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stochastic Models for Solar Power

Dimitra Politaki[1] and Sara Alouf[2]
[1]Université Côte d'Azur, Inria, CNRS, I3S, France
Dimitra.Politaki@inria.fr

[2]Université Côte d'Azur, Inria, France
Sara.Alouf@inria.fr

Author version

Abstract

In this work we develop a stochastic model for the solar power at the surface of the earth. We combine a deterministic model of the clear sky irradiance with a stochastic model for the so-called clear sky index to obtain a stochastic model for the actual irradiance hitting the surface of the earth. Our clear sky index model is a 4-state semi-Markov process where state durations and clear sky index values in each state have phase-type distributions. We use per-minute solar irradiance data to tune the model, hence we are able to capture small time scales fluctuations. We compare our model with the on-off power source model developed by Miozzo et al. (2014) for the power generated by photovoltaic panels, and to a modified version that we propose. In our on-off model the output current is frequently resampled instead of being a constant during the duration of the "on" state. Computing the autocorrelation functions for all proposed models, we find that the irradiance model surpasses the on-off models and it is able to capture the multiscale correlations that are inherently present in the solar irradiance. The power spectrum density of generated trajectories matches closely that of measurements. We believe our irradiance model can be used not only in the mathematical analysis of energy harvesting systems but also in their simulation.
keywords: Solar power · Semi-Markov process · Photovoltaic panel

1 Introduction

In the past decade, there has been an awareness rising concerning the energy cost and environmental footprint of the fastly growing Information and Communication Technology (ICT) sector. In [17] Van Heddeghem et al. assess how did the electricity consumption of the ICT sector evolve between 2007 and 2012. They report an increase in the relative share of ICT products and services (communication networks, personal computers and data centers, excluding TVs' set-top boxes and (smart)phones) in the total worldwide electricity consumption from about 3.9% in 2007 to 4.6% in 2012. Even though devices from new technologies are more energy efficient, this is outweighed by the fast growth in their numbers.

Among the most promising approaches recently pursued to reduce the environmental footprint of the ICT sector, we focus on the use of renewable energy sources and in particular solar energy. As photovoltaic panels are being used worldwide to power multiple components of the ICT sector, there is an increasing effort in the literature to consider the solar energy production when modeling computer and communication systems. For illustration purposes, we mention two recent papers modeling ICT systems involving renewable energy sources.

In [4], Dimitriou, Alouf and Jean-Marie consider a base station that is powered by renewable energy sources and evaluate in particular the depletion probability. The base station is modeled as a multi-queue queueing system where energy queues model the batteries that store the harvested energy. The authors of [4] model the renewable energy production as a Poisson process whose rate is modulated by a Markov chain representing the random environment.

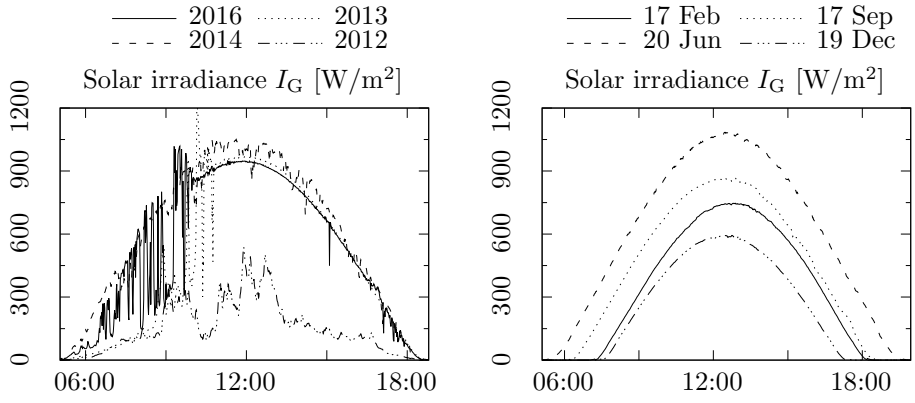
Neglia, Sereno and Bianchi consider in [13] the problem of geographical load balancing across data centers that have a dual power supply: grid and solar panels. They study the problem of scheduling jobs giving priority to data centers where renewable energy is available. The renewable energy source at each data center is modeled as an on-off process governed by a continuous time Markov chain. In the "on" state the data center can be fully powered by its renewable energy source; in the "off" state the data center is powered by the grid.

These examples among others illustrate the lack of a unified stochastic model for the solar energy to be used in the mathematical analysis of communication/computer systems. Our objective in this work is to develop such stochastic models for the solar power at the surface of the earth. Although there are a few models in the recent literature of the networking community [12], these rely on per-hour measurements. Therefore, such models do not capture the fluctuations in the solar irradiance at smaller time scales.

Our main contribution combines a deterministic model of the *clear sky* irradiance with a stochastic model of the so-called *clear sky index* to obtain a model of the actual irradiance hitting the surface of the earth. We will compare our model (after converting the actual irradiance to power generated by photovoltaic panels) to the night-day clustering model developed by Miozzo et al. in [12] for the generated power. We will propose for the latter a modified night-day clustering model. Our model for the harvested power is that of an on-off source in which the power generated in each state is frequently resampled from an appropriate distribution capturing the short-time scale fluctuations observed in practice.

To evaluate our models, we consider the autocorrelation functions and the periodograms of the generated trajectories. The autocorrelation function illustrates how well do our proposed models capture the multiscale correlations found in the data, whereas the spectral analysis allows to determine which characteristic time-scales are reproduced by the models.

In the following, we review the main notions used in the paper in Section 2 and discuss the related work in Section 3. Section 4 discusses the model of the clear sky index, and Section 5 is devoted to the model of the generated power. We assess our models in Section 6 before concluding the paper in Section 7.



(a) May 1st, different years

(b) Different days in 2012

Figure 1: Variations in the daily pattern of the solar irradiance are due to (a) the weather conditions and (b) the day of the year

2 Problem Definition

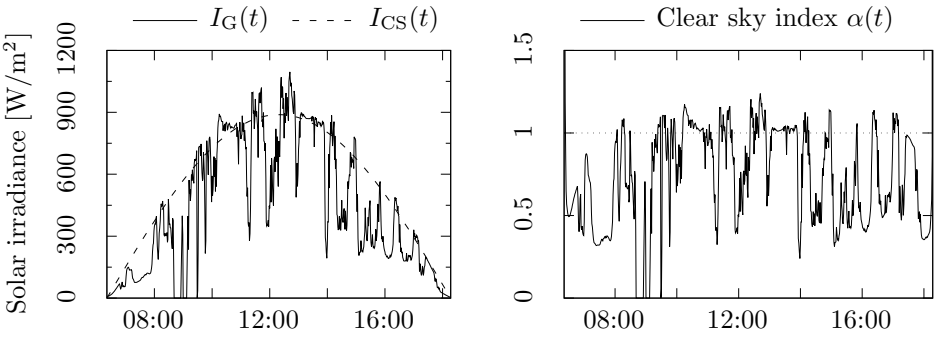
In this work, we are interested in two stochastic processes: the first one is the solar irradiance hitting a given surface on the earth, the second one is the power generated by a photovoltaic (PV) panel. We will define precisely each one of these processes in the following sections.

2.1 The Solar Irradiance

The amount of the solar energy that arrives per unit of time at a specific area of a surface is the solar irradiance and is expressed in W/m^2. In the following, the solar irradiance will refer to the *global* irradiance $I_G(t)$ accounting for all radiations arriving at a surface except for the ground-reflected ones. The reason for this is that we will rely on daily measurements of the global irradiance [1] to tune our models. No measurements of the ground-reflected radiations are available for download from [1]. However, their corresponding irradiance is usually insignificant compared to direct and diffuse irradiance.

The solar irradiance exhibits a night-day pattern that is affected by weather conditions which may induce burstiness at multiple time scales. Beside the obvious dependency on the geographic location, the solar irradiance depends also on the day of the year. Figure 1 illustrates these variations: per-minute measurements of the solar irradiance in Los Angeles [1] are depicted for the same day of different years (Fig. 1a) and for different days of the same year (Fig. 1b).

The solar irradiance $I_G(t)$ can be seen as the result of applying a multiplicative noise to the *clear sky* solar irradiance $I_{CS}(t)$. This multiplicative noise, denoted $\alpha(t)$ in this paper and called *clear sky index* in the literature, captures the perturbations seen in the solar irradiance with respect to the clear sky solar irradiance. We have $I_G(t) = \alpha(t) I_{CS}(t)$.



(a) Global and clear sky irradiances (b) Clear sky index $\alpha(t) = I_G(t)/I_{CS}(t)$

Figure 2: Illustrating the global irradiance $I_G(t)$, the clear sky irradiance model $I_{CS}(t)$ given in Eq. (2) and the resulting clear sky index $\alpha(t)$ on September 28th, 2010, in Phoenix, Arizona [16]

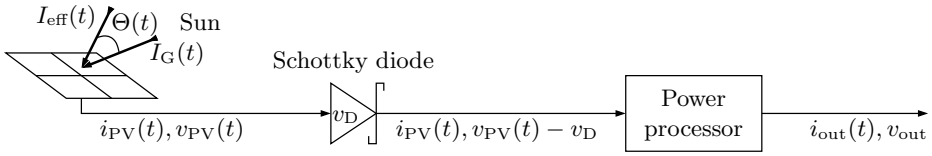


Figure 3: Using a fraction I_{eff} of the solar irradiance I_G, the PV cells generate a power (current i_{PV} and voltage v_{PV}) that goes through a Schottky diode and a power processor before it can be consumed

Figure 2 illustrates $I_G(t)$, $I_{CS}(t)$ and $\alpha(t)$ for a sample day.

2.2 The Power Generated by a PV Panel

The solar irradiance can yield electricity through the use of a PV panel as shown in Fig. 3. The usable power is directly related to the solar irradiance arriving at the panel (that is I_G) as thoroughly explained in [12] and implemented in the tool SolarStat that is available online [6]. The general idea is the following:

1. The solar irradiance effectively used by the PV panel is the component of $I_G(t)$ that is perpendicular to its surface, that is $I_{eff}(t)$.
2. The PV panel translates the effective solar irradiance I_{eff} into electric power with current $i_{PV}(t)$ and voltage $v_{PV}(t)$.
3. A Schottky diode reduces slightly the voltage but preserves the current.

4. A power processor extracts the maximum power from the PV panel and the output power has current $i_{\text{out}}(t)$ and voltage v_{out}.

The fluctuations seen in the solar irradiance $I_{\text{G}}(t)$ are still present in the output current $i_{\text{out}}(t)$. There may be additional fluctuations due to the local temperature and humidity that affect the functioning of the PV cells.

3 Related Work

Studies on the solar irradiance are abundant in the literature. Given the paramount role of the solar energy in many biological ecosystems, it is crucial to have models for the solar irradiance as measurements are not always available. For instance, Piedallu and Gégout develop in [14] a model that can predict the accumulated solar energy anywhere, providing annual figures for an entire country, as would be required for predictive vegetation modeling at a large scale. However such biology-oriented models are not fit for ICT applications that evolve typically on a much smaller time scales than vegetation.

Targeting the design of a solar system, there is a large body of work focusing on the clear sky irradiance. To cite a few references, Dave, Halpern and Myers overview in [3] several clear sky irradiance models and compare the accumulated daily and annual energy. They consider a tilted surface and account for both sky radiations and ground-reflected radiations. They find in particular that the effective irradiance at a surface is proportional to the cosine of the angle between the sunlight direction and the normal to the surface. Bird and Hulstrom compare in [2] five models for the maximum clear sky solar irradiance and propose a sixth model based on algebraic expressions. All these models require many meteorological input parameters (e.g., the surface pressure, the total ozone, the precipitable water vapor).

Another important component when modeling the solar irradiance is the clear sky index. Jurado, Caridad and Ruiz characterize the clear sky index using 5-minute measurements of the solar irradiance [10]. They partition the data according to the solar angle, considering two one-hour intervals at a time (both intervals corresponding to the same range of solar angle). They find that the density of the clear sky index in each partition is bimodal and can be modeled as a mixture of Gaussian distributions. The parameters of the distributions and the mixing factor are obtained from measurements by least squares approximation. The authors observe that the standard deviations of the Gaussian distributions depend on the solar angle. Also the bimodal behavior observed over 5-minute intervals is no longer observed when the interval in the data is larger. This is an important outcome that indicates that a model tuned with data having a given frequency of measurements can not match data having a different measurements rate. This observation supports our intuition that if one wants to use a model of solar power at a given time scale, then the model must be tuned with data at the same time scale. The authors of [10] are not clear on how do they compute the clear sky index from the measurements of the solar irradiance. Surprisingly, the computed clear sky index is always below 1 suggesting that they consider a very large maximum clear sky irradiance.

Gu et al. consider in [7] a related metric which is the relative change of solar irradiance (this would be $100(\alpha - 1)$) under the impact of clouds. They analyze per-minute measurements of solar irradiance collected in Brazil over a period of two months during

the wet season. They observe that broken cloud fields create a bimodal distribution for the relative change: shaded areas receive attenuated solar irradiance while sunlit areas may receive higher irradiance than under a clear sky. This effect is caused by radiations scattering and reflections from neighboring clouds. Conducting a spectral analysis on the time series of measured surface irradiance, they observe that clouds are responsible for two different regimes according to their types and density causing either large or small scale fluctuations. This study highlights the effect of clouds and have certainly impacted the development of subsequent models for the solar irradiance.

Miozzo et al. focus on the solar power generated by small embedded photovoltaic panels such as those used in sensor networks. They develop in [12] two stochastic models in which the dynamics of the power source is described by a semi-Markov process with $N \geq 2$ states. The first model is an on-off power source and the authors tune the sojourn time and power in each state by using a night-day clustering on hourly measurements of the solar irradiance. In the second model, the power source goes through a number of N states in a round-robin way and all sojourn times are equal and constant. A time slot based clustering enables the authors to estimate the power distribution in each state.

Ghiassi-Farrokhfal et al. consider also the solar power generated by photovoltaic panels but in the context of dimensioning an energy storage system. To near-optimally size a storage system, they develop in [5] a new *envelope* model for the generated power. In the general envelope model, the solar power is characterized by a statistical sample path lower envelope such that the probability of having the maximum of the distance envelope-solar energy exceed a given value is upper bound by a characteristic bounding function evaluated at the given value. Inspired by the findings of [7], the authors of [5] adapt the general envelope model to enable a separate characterization of the three underlying processes of solar power (diurnal, long-term, and short-term variations).

4 Modeling the Solar Irradiance I_{G}

In this section, we focus on the solar irradiance $I_{\mathrm{G}}(t)$. Our aim is to define a model able to capture the small time-scale fluctuations inherently present in the global irradiance. To that end, we model separately the clear sky irradiance $I_{\mathrm{CS}}(t)$ and the clear sky index $\alpha(t)$. By definition, we have

$$I_{\mathrm{G}}(t) = \alpha(t) I_{\mathrm{CS}}(t) \ . \tag{1}$$

We discuss $I_{\mathrm{CS}}(t)$ in Section 4.1 and model $\alpha(t)$ in Section 4.2.

4.1 Modeling the clear sky irradiance $I_{\mathrm{CS}}(t)$

The solar irradiance arriving at a surface during a clear sky day without any perturbations due to a change in the meteorological conditions exhibits a predictable pattern as shown in Fig. 1b. The models discussed in [3] for the hourly clear sky irradiance and in [2] for the maximum clear sky irradiance are not easily applicable given the unavailability of many input parameters. Instead, we use the so-called "simple sky model" [9] which defines a simple sinusoidal form for each day, taking into account the times of sunrise and sunset and the maximum clear sky irradiance. The clear sky irradiance $I_{\mathrm{CS}}(t)$ is given by the

following equation:

$$I_{\mathrm{CS}}(t) = \mathrm{MaxClearSky} \cdot \sin\left(\frac{t - \mathrm{sunrise}}{\mathrm{sunset} - \mathrm{sunrise}}\pi\right) \ . \tag{2}$$

The values of "sunrise", "sunset" and "MaxClearSky" are astronomical data that can be easily obtained in practice for any date and many selected locations from the website [15] (the maximum clear sky irradiance is called there "maximal solar flux"). An illustration of Eq. (2) is in Fig. 2a.

4.2 Modeling the clear sky index $\alpha(t)$

The clear sky index $\alpha(t)$ captures the fluctuations over time of the global irradiance with respect to clear sky conditions, as illustrated in Fig. 2b for a sample day and a sample location. Consequently, one thinks of defining a state for each macro weather condition. Based on our review of the literature, we define four states for $\alpha(t)$ that correspond to: heavy clouds between the sun and the surface (very low values of $\alpha(t)$), medium to light clouds between the sun and the surface (values of $\alpha(t)$ around 0.6), clear sky (values of $\alpha(t)$ around 1), and high reflection and diffusion in the atmosphere (values of $\alpha(t)$ larger than 1). We assume all transitions between different states to be possible.

We propose to capture the dynamics of $\alpha(t)$ by a discrete-time semi-Markov process.[1] Our model works as follows. When the process $\alpha(t)$ enters a state i, it will remain there for a duration τ_i governed by a probability density function f_i. While in state i, the clear sky index $\alpha(t)$ behaves like $\alpha_i(t)$, a stochastic process with probability density function g_i. When the sojourn time τ_i expires, the process *changes* its state. The distributions f_i and g_i, for $i \in \{1, 2, 3, 4\}$ will be fitted to empirical distributions of the sojourn times and values of $\alpha(t)$.

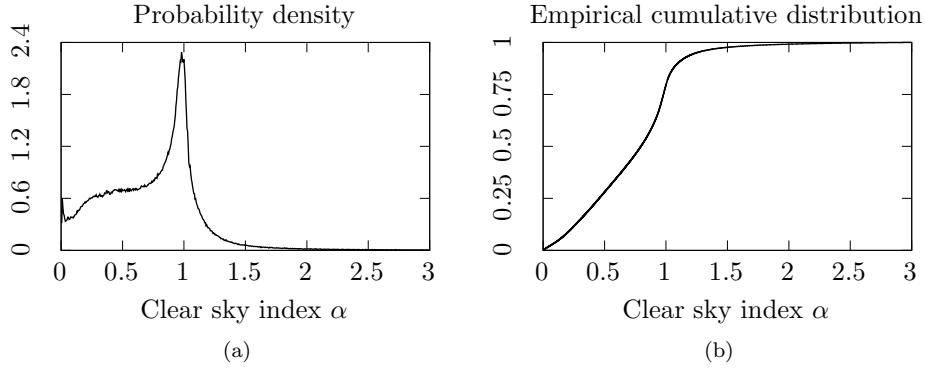
To tune our model of $\alpha(t)$ we use per-minute measurements of the solar irradiance $I_{\mathrm{G}}(t)$. The data is for the region of Los Angeles from April 2010 until March 2015 [1]. We compute $\alpha(t) = I_{\mathrm{G}}(t)/I_{\mathrm{CS}}(t)$ using the data and Eq. (2) for each minute during the five years.[2] For illustration purposes, we compute the density and the cumulative distribution of the clear sky index and depict them in Fig. 4.

Remark 1 *The density of the clear sky index depicted in Fig. 4a is not bimodal as found in [10]. The measurements used in [10] were made every 5 minutes and the densities were computed over two intervals of 1 hour each corresponding to the same range of the solar angle. Instead, the density shown in Fig. 4a is for all 1-minute measurements over a period of 5 years.*

Once that we have computed the values of $\alpha(t)$, we first aim to validate the number of states of our semi-Markov model. We apply the k-means clustering algorithm [11] and use the Davies-Bouldin index to define the optimal number of clusters. The Davies-Bouldin

[1] Using a discrete-time Markov process does not yield satisfactory results as correlations are not described well.

[2] We observe that we may well have in the real measurements $I_{\mathrm{G}}(t) > 0$ around sunset and sunrise due to diffusion. As $I_{\mathrm{CS}}(t) = 0$ at sunrise (and before) and sunset (and after), this implies that infinite values for the ratio $I_{\mathrm{G}}(t)/I_{\mathrm{CS}}(t)$ can occur. To discard such values when computing $\alpha(t)$, we enforce the (arbitrary) bound $\alpha(t) < 3$.



(a) (b)

Figure 4: Density and cumulative distribution curves of the clear sky index $\alpha(t)$ computed using Eq. (2) and per-minute solar irradiance data [1]

Table 1: Values in each cluster according to k-means, their corresponding state in the semi-Markov model and weather condition

Range of values of $\alpha(t)$	State	Physical interpretation
$[0, 0.44152)$	1	heavy clouds between the sun and the surface
$[0.44152, 0.81639)$	2	medium to light clouds between the sun and the surface
$[0.81639, 1.4343)$	3	clear sky
$[1.4343, 3)$	4	high reflection and diffusion in the atmosphere

index is based on a ratio of within-cluster and between-cluster distances. The smaller its value the better the clustering.

We tested nine different clustering (for $k \in \{2, \ldots, 10\}$) and computed the Davies-Bouldin index for each clustering obtained. The values of the index were between 0.5017 and 0.5290. The smallest value was obtained for $k = 4$ implying that ideally the values of $\alpha(t)$ should be classified into four clusters. This analysis supports our choice of having four states in the model for the clear sky index and each state is mapped to one of the four clusters obtained. The details on the four clusters/states obtained when applying the k-means clustering algorithm are given in Table 1.

Now that we have clearly identified the four states of our semi-Markov model, our next step is to identify the transition probabilities among the states. We estimate them using the computed values of $\alpha(t)$ and the identified clusters. We first map each computed value of $\alpha(t)$ to its corresponding state, then we count the number of transitions between any pair of states. The transition probability from state i to state j is estimated as the ratio of the number of transitions from state i to state j to the total number of transitions out of state i. We find the following transition probability matrix for the four-state semi-Markov

Table 2: Number of phases of the phase-type distribution fitting the (shifted) sojourn times and values in each state

Variable	Number of samples used in the fitting	Number of phases of the phase-type distribution fitting the variable
$\tau_1 - 1$	19678	5
$\tau_2 - 1$	8456	6
$\tau_3 - 1$	2094	6
$\tau_4 - 1$	15400	6
$\alpha_1(t)$	298141	20
$\alpha_2(t) - 0.44152$	345973	20
$\alpha_3(t) - 0.81639$	563411	6
$\alpha_4(t) - 1.4343$	34432	3

model:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0.8361 & 0.0549 & 0.1090 \\ 0.3645 & 0 & 0.6296 & 0.0059 \\ 0.0274 & 0.9019 & 0 & 0.0707 \\ 0.0484 & 0.0536 & 0.8980 & 0 \end{bmatrix} . \tag{3}$$

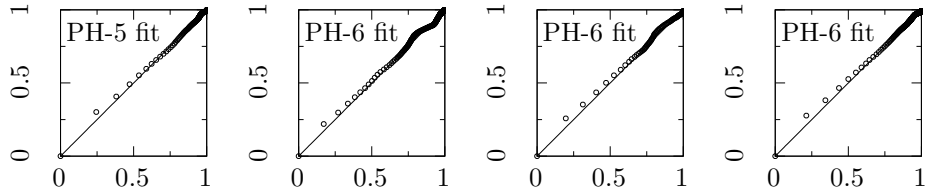
The last step is to characterize the densities f_i and g_i for $i = 1, \ldots, 4$. We carry out a statistical analysis on the computed values of $\alpha(t)$ in order to determine the distributions of the sojourn times $\{\tau_i\}_{i=1..4}$ and the values $\{\alpha_i(t)\}_{i=1..4}$. Observe that the sojourn time τ_i in a given state i corresponds to the number of consecutive values of $\alpha(t)$ inside the corresponding cluster. Recall that $\alpha(t)$ is a discrete-time process and as the measurements used for tuning the model are minute-based, therefore the time unit in our model is the minute.

We opt to fit the data with phase-type (PH) distributions given their attractive analytical tractability and their high flexibility in fitting data. We use the PhFit tool [8] to find the phase-type distribution that best fits each one of the empirical distributions. In the PhFit tool, we choose the relative entropy as distance measure according to which the fitting is performed.

We repeatedly fit the data related to each variable changing the number of phases. We use probability plots to assess the quality of the fit and select the number of phases that yields the best fit. We report the chosen number of phases for each fitted variable in Table 2.

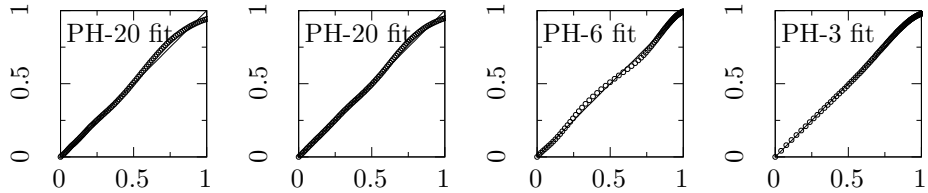
The probability plots of the selected phase-type distributions are displayed in Figures 5-6. Each graph in Fig. 5 depicts on the y-axis the probabilities of the fitted distribution against the probabilities of the sojourn times in a given state on the x-axis. We observe that the phase-type distribution fits reasonably well the sojourn times for all states.

Regarding the values of $\alpha(t)$ in each state, we can see in Fig. 6 that the selected phase-type distributions fit very well the values of $\alpha(t)$. We observe that the quality of fit for $\alpha_1(t)$ and $\alpha_2(t)$ is obtained at the cost of having a significantly larger number of phases (that is 20; see Table 2) with respect to the other variables.



(a) $\tau_1 - 1$ (b) $\tau_2 - 1$ (c) $\tau_3 - 1$ (d) $\tau_4 - 1$

Figure 5: Probability plots of the phase-type fitting for sojourn times in each state



(a) $\alpha_1(t)$ (b) $\alpha_2(t) - 0.44152$ (c) $\alpha_3(t) - 0.81639$ (d) $\alpha_4(t) - 1.4343$

Figure 6: Probability plots of the phase-type fitting for $\alpha(t)$ values in each state

5 Modeling the Harvested Power

To account for the power generated by PV panels when evaluating solar-powered systems, one has mainly two options. The first option is to use a model for the solar irradiance such as the one developed in Section 4 and then infer the power generated by the PV cells (or equivalently $i_{\text{out}}(t)$; see Fig. 3). This second step may be a simple linear model (i.e. the power generated by a panel of unit size is the solar irradiance effectively received multiplied by the efficiency of the panel) or a more detailed model such as the one implemented in the SolarStat tool [6]. The second option is to use directly a model for the power generated by a given PV panel (i.e. a model for i_{out}). Miozzo et al. have developed two such models in [12]. In this section, we propose a modification to their on-off model. We will compare our modified model to theirs in Section 6 and also to the model of Section 4 after we translate the solar irradiance to generated power using the SolarStat tool. We present briefly the on-off model in [12] before explaining our modification.

The dynamics of the harvested current $i_{\text{out}}(t)$ are captured by a two-state semi-Markov process. The distributions of the sojourn times and of $i_{\text{out}}(t)$ in each state are statistically defined using hourly measurements of the solar irradiance. In practice, Miozzo et al. apply the procedure summarized in Section 2.2 to map the solar irradiance data into the power generated by a PV panel of given size (number of solar cells connected in series/parallel) and characteristics (open circuit voltage, short circuit current, and reference temperature). Assuming the output voltage to be constant, the generated power and the output current are proportional to each other. The mapped data is grouped by month and for each month the values of the *output current* $i_{\text{out}}(t)$ are classified into two states according to an

arbitrarily low threshold. All points falling below the threshold correspond to the "night" state and points falling above the threshold correspond to the "day" state. The authors of [12] use kernel-smoothing techniques to estimate the distributions of the durations and output current in each state for every month of the year. The model is as follows: when entering a state, a current and a duration are drawn from the corresponding distributions, then the source outputs the drawn current *constantly* for the drawn duration. At the end of the drawn duration, the source switches its state. In practice, the output current in the night state is set to 0.

Modified On-Off Model. To better capture the fluctuations observed in the solar irradiance $I_{\mathrm{G}}(t)$ (which will inevitably be present in $i_{\mathrm{out}}(t)$), we propose to modify the above-mentioned model in the following way: instead of keeping the current *constant* during the time the process remains in the "day" state, we frequently *resample* (every ten minutes) from the current distribution until a transition occurs.

6 Results

In this section we will evaluate the models presented in Sections 4 and 5. We consider first the autocorrelation function (ACF) as a metric to test how well do generated synthetic data match the empirical data according to second order statistics. The empirical data is a 5-year long set of output current values sampled every minute. The current values are those matched by SolarStat (for a Panasonic solar panel of unit size) for the solar irradiance measurements [1]. We generate three synthetic data that are:

1. a 5-year long set of output current values sampled every minute using the model of the solar irradiance presented in Section 4 and SolarStat to translate the irradiance into output current;

2. a 5-year long set of output current values sampled every 10 minutes using the on-off model in [12];

3. a 5-year long set of output current values sampled every 10 minutes using our modified on-off model (Section 5).

The autocorrelation functions of these four data sets are depicted in Fig. 7. Our solar irradiance model performs fairly well, capturing most of the correlations present in the empirical data. As already found by the authors of [12], the ACF of the on-off source model poorly resembles that of the empirical data. The ACF of our modified on-off model performs seemingly equally badly.

Strong correlations in the solar power exists over yearly lags due to the earth's annual circumnavigation of the sun. To assess how well does our solar irradiance model capture the correlations over very long periods, we sample the ACFs every 30 days and display the values in Fig. 8. We can make three observations: first, the ACF of the real data confirms the expected strong annual correlation; second, our solar irradiance model exhibits correlations that mimic those in the real data, even though to a lesser extent; third, the on-off models fail to track the ACF of the real data.

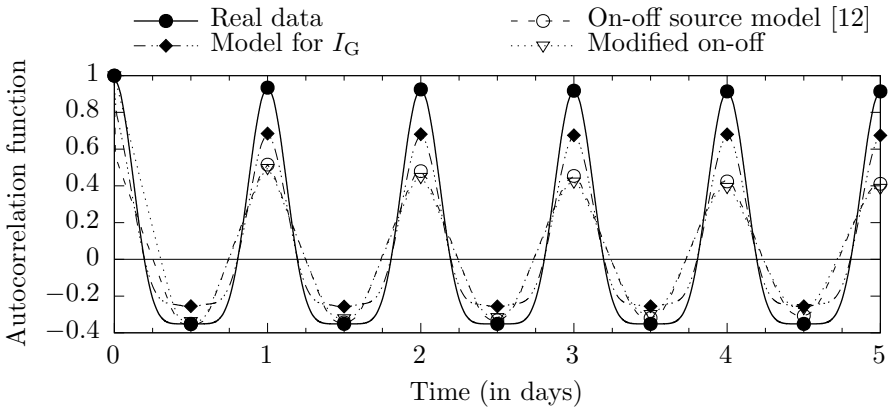


Figure 7: ACF of the current harvested using Panasonic solar panels

Table 3: Root mean square error (RMSE) between real and synthetic data

Model of solar	Model of harvested power	
irradiance I_G	On-off source model [12]	Modified on-off
0.1274	0.3231	0.2839

To complete this comparative analysis of the models, we compute the root mean square error (RMSE) between the ACF of the empirical data set and that of each of the synthetic data set. The RMSE metric is as follows: $\text{RMSE} = \sqrt{\frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2}$, where y_i and $\hat{y}_i$ are the ith samples of the empirical and synthetic data respectively, and n is the number of samples. The results reported in Table 3 confirm the superiority of the solar irradiance model over the on-off models.

We can conclude from the comparison of the ACFs that our model of the solar irradiance outperforms the on-off models of the output current and captures well the multiscale correlations found in the real data.

We consider next the periodograms of the empirical data set and the synthetic data set generated by the solar irradiance model (see Section 4). The spectral analysis allows to determine which characteristic time-scales are reproduced by the model.

We compute the periodogram using the function with the same name in the Signal Processing Toolbox of Matlab. We adjust appropriately the x-axis in order to have frequencies (f, in Hertz) instead of the angular frequency ω. The power spectrum densities (PSD) are depicted in Figs. 9 and 10.

Observe that Gu et al. have analyzed in [7] the power spectrum of a 2-month set of 1-minute measurements of solar irradiance. The PSD had two clear peaks corresponding to 24 and 12 hours but other than those the absence of other characteristic time-scale was striking. This is not the case of the PSD of the real data set displayed in Fig. 9. We can observe a series of peaks at larger frequencies that are the harmonics of $1.157\overline{407}\ 10^{-5}$ Hz (which corresponds to 24 hours). The same observation applies to the PSD of the synthetic data set displayed in Fig. 10. The peak at the fundamental frequency corresponding to 1

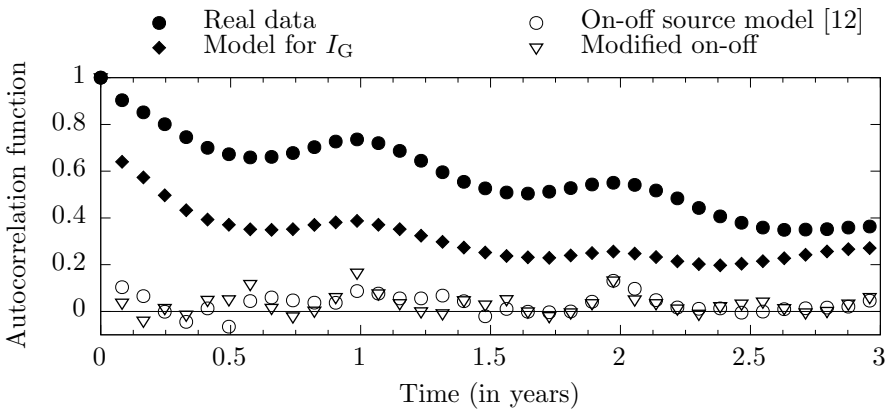


Figure 8: Samples of the autocorrelation function of the output current (one sample per 30 days)

day is clearly visible as well as those of its harmonics frequencies.

We conclude this section by stating that our solar irradiance model is able to generate synthetic data that exhibits all of the frequency peaks of real data, capturing its characteristic time-scales.

7 Conclusions

We have developed in this work a stochastic model for the solar irradiance. The model combines a deterministic model of the clear sky irradiance with a stochastic model for the so-called clear sky index to obtain a stochastic model for the actual irradiance hitting the surface of the earth. As per-minute solar irradiance data is used to tune our model, we are able to capture small time scales fluctuations as would be needed by ICT applications. Computing autocorrelation functions and periodograms of empirical and synthetic traces we found that our solar irradiance model performs very well. We believe our model can be used not only in the mathematical analysis of energy harvesting communication/computer systems but also in their simulation.

8 Acknowledgements

The authors would like to thank Alain Jean-Marie for fruitful discussions during early stages of this work. This work was partly funded by the French Government (National Research Agency, ANR) through the "Investments for the Future" Program reference #ANR-11-LABX-0031-01.

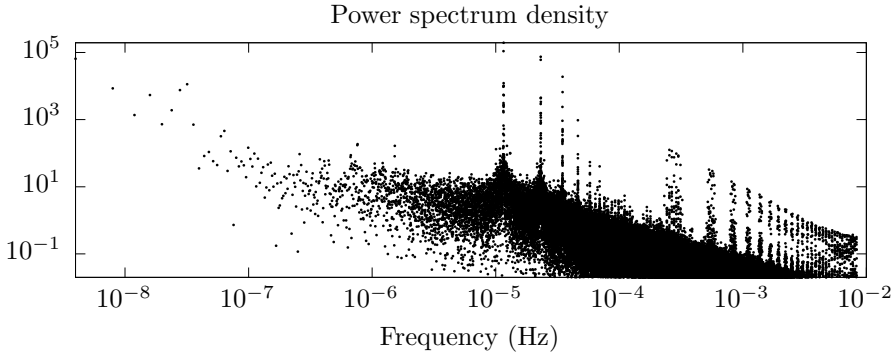


Figure 9: Power spectrum of the 1-minute values of output current mapped from the real measurements [1] by the SolarStat tool

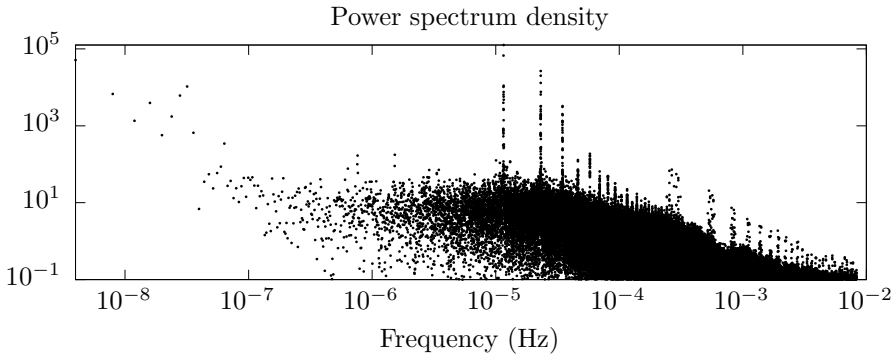


Figure 10: Power spectrum of the 1-minute values of output current obtained after generating a 5-year trajectory from the model of Section 4 and translating it to current with the SolarStat tool

References

[1] A. Andreas and S. Wilcox. Solar Resource & Meteorological Assessment Project (SOLRMAP): Rotating Shadowband Radiometer (RSR); Los Angeles, California (Data). Report DA-5500-56502, NREL, 2012.

[2] Richard E. Bird and Roland L. Hulstrom. A simplified clear sky model for direct and diffuse insolation on horizontal surfaces. Technical Report Technical report SERI/TR-642-761, Solar Energy Research Institute, February 1981.

[3] J. V. Dave, P. Halpern, and H. J. Myers. Computation of incident solar energy. *IBM Journal of Research and Development*, 19(6):539–549, November 1975.

[4] Ioannis Dimitriou, Sara Alouf, and Alain Jean-Marie. A markovian queueing system for modeling a smart green base station. In *Proceedings of EPEW: European Per-*

formance Evaluation Workshop, volume 9272 of *LNCS*, pages 3–18, Madrid, Spain, August 2015.

[5] Yashar Ghiassi-Farrokhfal, Srinivasan Keshav, Catherine Rosenberg, and Florin Ciucu. Solar power shaping: An analytical approach. *IEEE Transactions on Sustainable Energy*, 6(1):162–170, January 2015.

[6] Marco Gianfreda, Marco Miozzo, and Michele Rossi. SolarStat: Modeling photovoltaic sources through stochastic Markov processes. Online. http://www.dei.unipd.it/∼rossi/Software/Sensors/SolarStat.zip.

[7] Lianhong Gu, Jose D. Fuentes, Michael Garstang, Julio Tota da Silva, Ryan Heitz, Jeff Sigler, and Herman H. Shugart. Cloud modulation of surface solar irradiance at a pasture site in southern Brazil. *Agricultural and Forest Meteorology*, 106(2):117–129, January 2001.

[8] András Horváth and Miklós Telek. PhFit: A general phase-type fitting tool. In *Proceedings of TOOLS: International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, volume 2324 of *LNCS*, pages 82–91, London, UK, April 2002.

[9] Muhammad Iqbal. *An Introduction to Solar Radiation*. Academic Press, 1983.

[10] M. Jurado, J.M. Caridad, and V. Ruiz. Statistical distribution of the clearness index with radiation data integrated over five minute intervals. *Solar Energy*, 55(6):469–473, December 1995.

[11] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):881–892, July 2002.

[12] Marco Miozzo, Davide Zordan, Paolo Dini, and Michele Rossi. SolarStat: Modeling photovoltaic sources through stochastic Markov processes. In *Proc. of 2014 IEEE International Energy Conference*, pages 688–695, Dubrovnik, Croatia, May 2014.

[13] Giovanni Neglia, Matteo Sereno, and Giuseppe Bianchi. Geographical Load Balancing across Green Datacenters. *ACM SIGMETRICS Performance Evaluation Review*, 44(2):64–69, September 2016.

[14] Christian Piedallu and Jean-Claude Gégout. Multiscale computation of solar radiation for predictive vegetation modelling. *Annals of Forest Science*, 64(8):899–909, January 2007.

[15] ptaff.ca. Sunrise, sunset daylight in a graph.

[16] Solar Resource & Meteorological Assessment Project (SOLRMAP), Southwest Solar Research Park (Formerly SolarCAT).

[17] Ward Van Heddeghem, Sofie Lambert, Bart Lannoo, Didier Colle, Mario Pickavet, and Piet Demeester. Trends in worldwide ICT electricity consumption from 2007 to 2012. *Computer Communications*, 50:64–76, September 2014.